Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

PETER DEUFLHARD

Adaptive Pseudo–transient Continuation for Nonlinear Steady State Problems

ZIB-Report 02–14 (March 2002)

Adaptive Pseudo–transient Continuation for Nonlinear Steady State Problems

Peter Deuflhard

Abstract

Pseudo–transient continuation methods are quite popular for the numerical solution of steady state problems, typically in PDEs. They are based on an embedding into a time dependent initial value problem. In the presence of dynamical invariants the Jacobian matrix of the nonlinear equation system is bound to be singular. The paper presents a convergence analysis which takes this property into account – in contrast to known approaches. On the basis of the new analysis adaptive algorithms are suggested in detail. These include a variant with Jacobian approximations as well as inexact pseudo–transient continuation, both of which play an important role in discretized PDEs. Numerical experiments are left to future work.

Keywords: pseudo–transient continuation, linearly implicit Euler discretization, stiff integration, contractivity of ordinary differential equations, large scale nonlinear systems, discretized partial differential equations, inexact Newton techniques, affine invariance

MSC (1991): 65H10, 65H20, 65L05, 65N06, 65N22

Contents

1 Introduction

This paper deals with the numerical solution of nonlinear systems

$$F(x) = 0$$

which can be interpreted as *steady state problems* of a corresponding *dynamical system*

$$\dot{x} = F(x), \quad x(0) = x_0 \; . \tag{1.1}$$

The problem class mainly under consideration here are steady state problems originating from discretized partial differential equations.

From mere geometrical insight it is already clear that any such approach can only work, if the fixed point of the dynamical system is *attractive*. If the fixed point were *hyperbolic* (in the sense of dynamical systems, not of PDEs !), then numerical integration of (1.1) would generically come close to the fixed point in an initial phase, but "run away" afterwards; in such a case, convergence would only occur from exceptional starting points x_0 out of a set of measure zero. The basic idea pursued here is to compute attractive fixed points by a special technique borrowed from stiff integration. Among the many possibilities attention is focused on the *linearly implicit Euler discretization* – compare, e.g., the textbooks by Hairer and Wanner [5] or by Deuflhard and Bornemann [3].

In the context of solving nonlinear systems this discretization is often called *pseudo–transient continuation*. Recently, Kelley and Keyes [6] gave a convergence analysis of the method in view of an adaptive realization of the algorithm. This analysis required nonsingularity of the Jacobian matrix. However, dynamical invariants, which may be even hidden, will typically cause a non-empty nullspace of the Jacobian. The topic of this paper is to present an alternative convergence analysis, which does note require the nonsingularity of the Jacobian and explicitly allows for the possible occurrence of (possibly unknown) dynamical invariants. In Section 2 we collect preliminary considerations concerning affine invariance aspects, the role of dynamical invariants, and the characterization of attractive fixed points in the linear case. On this basis, Section 3 presents our convergence analysis for the nonlinear case in terms of carefully chosen first and second order Lipschitz conditions. This analysis is then exploited for the construction of an adaptive pseudo–transient continuation method with exact Jacobian and direct elimination method for the arising linear systems. The practically relevant

cases of an approximate Jacobian or of an iterative solution of the arising linear systems is then treated in a synoptic manner in Sections 4 and 5. For both situations adaptive algorithms are suggested. Numerical experiments with the suggested algorithms are not given here, but left to future work.

2 Preliminary considerations

In this section we first want to gain some preliminary insight into the class of promising algorithms for the solution of problem (1.1). For this reason we perform a linear coordinate transformation, in terms of an arbitrary nonsingular matrix B, such that

$$x \longrightarrow y = Bx \ .$$

Insertion into (1.1) yields the transformed problem

$$\dot{y} = BF(B^{-1}y) =: G(y) \ , \quad y(0) = Bx_0 \ . \tag{2.1}$$

Whenever the original problem $F(x) = 0$ has the unique solution x, then the transformed problem

$$G(y) = BF(B^{-1}y) = 0 \tag{2.2}$$

has the unique solution $y = Bx$. The effect on the corresponding Jacobian is a similarity transformation

$$G'(y) = BF'(x)B^{-1} \ .$$

That is why we say that the problem classes (2.1) and (2.2) have the property of *affine similarity* which essentially states that the domain and the image space transform in the same way. Of course, we will try to observe this property both in the construction of algorithms and in their theoretical analysis.

As can be easily seen, affine similarity allows for Newton iterations like

$$F'(x^k)\Delta x^k = -F(x^k), \quad x^{k+1} = x^k + \lambda_k \Delta x^k, \qquad k = 0, 1, \ldots$$

and for fixed point iterations

$$\Delta x^k = F(x^k), \quad x^{k+1} = x^k + \lambda_k \Delta x^k, \qquad k = 0, 1, \ldots$$

with a parameter λ to be adapted at each iterative step. The same holds, of course, for any linear combination of Newton and fixed point iteration such as the popular so–called *pseudo–transient continuation method*

$$(I - \tau A)\,\Delta x = F(x_0), \quad x(\tau) = x_0 + \tau \Delta x \tag{2.3}$$

with pseudo–timestep τ to be adapted. Herein the matrix A is either the Jacobian $F'(x_0)$ or an appropriate approximation. The above scheme is also known as the *linearly implicit Euler discretization* of the time dependent problem (1.1) – see, e.g., [5, 3].

Dynamical invariants. In the class of problems under consideration, dynamical invariants occur rather frequently. As an example, take *mass conservation*, which typically shows up as

$$e^T x(t) = e^T x_0 \ ,$$

where $e^T = (1, \ldots, 1)$. This implies, for arbitrary $F(x) \neq 0$, that

$$e^T \dot{x} = e^T F(x) = 0 \quad \text{for all} \quad x \in D \subset \mathbb{R}^n \ .$$

By differentiation with respect to time t we obtain

$$e^T F'(x) F(x) = 0 \quad \text{for all} \quad x \in D \subset \mathbb{R}^n$$

and hence the Jacobian (at all arguments) has a zero eigenvalue with left eigenvector e. If we define the orthogonal projectors

$$P^\perp := \frac{1}{n} e e^T, \quad P = I - P^\perp \ ,$$

then we can write equivalently

$$P^\perp F'(x) = 0 \ .$$

As a consequence, naive application of any standard Newton method would fail in this situation. For this a–priori known dynamical invariant a modification of Newton's method is possible such as

$$F'(x^k)\Delta x^k = -F(x^k), \quad e^T \Delta x^k = 0 \ .$$

In general, however, more dynamical invariants may exist, most of them either unspecified or even unknown. Thus, in terms of some projector P we again have

$$P^\perp \dot{x} = P^\perp F(x) = 0 \quad \Longrightarrow \quad P^\perp F'(x) = 0 \ .$$

This time, however, the projector is unknown so that Newton's method cannot be successfully modified and is therefore bound to fail.

Upon examining the above iterations, we easily verify that the fixed point iteration as well as the pseudo–transient iteration do not require any nonsingularity of the Jacobian: both automatically realize the condition

$$P^{\perp}\Delta x = 0 \; .$$

In other words: the desired dynamical invariance property is implicitly inherited to these iterations even in the case of an unknown projector P.

Linear Contractivity. In order to find a theoretical characterization for *attractive* fixed points, we consider, for the time being, linear problems

$$\dot{x} = Ax, \; x(0) = x_0 \; , \tag{2.4}$$

which give rise to a formal solution in terms of the matrix exponential

$$x(t) = \exp(At)x_0 \; .$$

Any coordinate transformation $y = Bx$ would imply a similarity transformation on the matrix

$$A \longrightarrow BAB^{-1} \; .$$

In view of this behavior, we may start from the Jordan decomposition

$$A = TJT^{-1} \longrightarrow \big(BT\big)J\big(BT\big)^{-1} \; .$$

Herein J is the *Jordan canonical form* consisting of elementary Jordan blocks for each separate eigenvalue $\lambda(A)$. Then the transformation

$$z := T^{-1}(x - \hat{x}) = \big(BT\big)^{-1}(y - \hat{y})$$

with origin $\hat{y} = B\hat{x}$ generates an affine similar coordinate frame.

In what follows we will have to work with norms $\|\cdot\|$ induced by certain inner products $(\cdot,\cdot)$. For simplicity, we may think of the Euclidean norm $\|\cdot\|$ induced by the Euclidean inner product $(u,v) = u^T v$ or any *scaled* variants of them (preconditioning). If we formulate our subsequent theoretical statements in terms of the *canonical norm*

$$|u| := \|T^{-1}u\| \; ,$$

induced by the *canonical inner product*

$$\langle u, v\rangle = (T^{-1}u, T^{-1}v) \ ,$$

then such statements will automatically meet the requirement of affine similarity. In this setting, we may define some constant $\mu = \mu(A)$, allowed to be *positive, zero,* or *negative*, such that

$$\langle u, Au\rangle \leq \mu(A)|u|^2 \ . \tag{2.5}$$

This definition is obviously equivalent to

$$(\bar{u}, J\bar{u}) \leq \mu\|\bar{u}\|^2 \ , \tag{2.6}$$

wherein $\bar{u} = T^{-1}u$. Assuming that the quantity μ is chosen best possible, it can be shown to satisfy

$$\mu(A) = \max_{u\neq 0} \frac{\langle u, Au\rangle}{|u|^2} \geq \max_i \Re\lambda_i(A) + \epsilon \ , \quad \epsilon \geq 0 \ . \tag{2.7}$$

Herein $\epsilon = 0$ holds, if the eigenvalue of A with dominating real part is simple. By the above affine transformation we can show that

$$\mu(BAB^{-1}) = \mu(A) \ , \tag{2.8}$$

which confirms that this quantity is affine similar. We may apply it to obtain the estimate

$$|x(t)| \leq \exp(\mu t)|x_0| \ .$$

Whenever

$$\mu < 0$$

holds, then

$$|x(t)| < |x_0| \ ,$$

which means that the linear dynamical system (2.4) is *contractive*.

For computational reasons, the Euclidean product (possibly in a scaled variant) is favored over the canonical product. Suppose that we therefore replace the above definition (2.5) by the analogous definition

$$\nu(A) = \max_{u\neq 0} \frac{(u, Au)}{\|u\|^2} \tag{2.9}$$

in terms of the Euclidean product. The thus defined quantity can be expressed as (cf. [4, Sect. 1.10])

$$\nu(A) = \lambda_{\max}\left(\frac{1}{2}(A + A^T)\right) ,$$

where $\lambda_{\max}$ is the maximum (real) eigenvalue of the symmetric part of the matrix A. On one hand, by comparison with (2.6), we may observe that

$$\mu(A) = \nu(J) = \lambda_{\max}\left(\frac{1}{2}(J + J^T)\right) ,$$

which directly leads to the above result (2.7). On the other hand, the quantities $\nu(A)$ and $\mu(A)$ may be rather different. In general, for arbitrary transformation matrix B, we will – in contrast to (2.8) – obtain

$$\nu(BAB^{-1}) \neq \nu(A) , \tag{2.10}$$

i.e. the quantity ν is not affine similar. Not even the signs of $\nu(A)$ and $\mu(A)$ need to be the same.

As a consequence of this structure, contractivity of the differential equation (1.1) will show up as contractivity in the canonical norm $|\cdot|$, but need not show up as contractivity in the Euclidean norm $\|\cdot\|$. In fact, whenever

$$\nu < 0 \tag{2.11}$$

holds, then

$$\|x(t)\| \leq \exp(\nu t)\|x_0\| < \|x_0\| .$$

Generally speaking, if we transform a relation of the kind

$$|u| \leq |v| ,$$

to the Euclidean norm, then we can only prove that

$$\|u\| \leq \mathrm{cond}(T)\|v\|$$

in terms of the condition number $\mathrm{cond}(T) = \|T^{-1}\| \cdot \|T\| \geq 1$. The condition number here arises as an unavoidable *geometric distortion* factor which indicates the possible *ill–conditioning* of the Jordan decomposition as a whole.

3 Exact pseudo–transient continuation

We now want to study the scheme (2.3) as an iterative method. Throughout this section we assume that we can evaluate an *exact Jacobian* $A = F'(x)$ and solve the linear system (2.3) by *direct* elimination techniques. Mollifications of these assumptions will be treated in subsequent sections.

Let $x(\tau)$ denote the *homotopy path* defined by (2.3) starting at $x(0) = x_0$. Before we actually study the residual norm $\|F(x(\tau))\|$, the following auxiliary result will be helpful. It holds for exact as well as for approximate Jacobian A.

Lemma 3.1 *Notation as just introduced with $A \approx F'(x_0)$. Then the residual along the homotopy path $x(\tau)$ starting at x_0 satisfies*

$$F(x(\tau)) = (I - \tau A)^{-1} F(x_0) + \int_{\sigma=0}^{\tau} (F'(x(\sigma)) - A)\,(I - \sigma A)^{-2} F(x_0) d\sigma \ . \quad (3.1)$$

Proof. Taylor's expansion of the residual yields

$$\begin{aligned} F(x(\tau)) &= F(x_0) + \int_{\sigma=0}^{\tau} F'(x(\sigma))\dot{x}(\sigma) d\sigma \\ &= F(x_0) + (x(\tau) - x_0) + \int_{\sigma=0}^{\tau} (F'(x(\sigma)) - A)\,\dot{x}(\sigma) d\sigma \ . \end{aligned}$$

Upon differentiating the homotopy (2.3) with respect to τ, we obtain

$$(I - \tau A)\dot{x} = F(x_0) + A(x(\tau) - x_0) = F(x_0) + \tau A (I - \tau A)^{-1} F(x_0)$$

and therefore

$$\dot{x}(\tau) = (I - \tau A)^{-2} F(x_0) \ ,$$

which then readily leads to the result of the lemma. ∎

Discussion of Lipschitz conditions. With the above auxiliary result at hand, the question is now how to formulate first and second order Lipschitz conditions in view of a theoretical convergence analysis.

First order Lipschitz condition. We will employ some *one–sided Lipschitz constant* ν as introduced in the preceding section. Due to the possible occurrence of dynamical invariants the Jacobian matrix may have zero eigenvalues, which would imply $\nu \geq 0$ instead of the required $\nu < 0$. Therefore we will restrict our attention to iterative corrections in the subspace

$$S_P = \{u \in \mathbb{R}^n \mid P^{\perp} u = 0\} \; .$$

Then the definition inequality

$$(u, Au) \leq \nu \|u\|^2, \quad u \in S_P$$

is equivalent to

$$(Pu, (PAP)Pu) \leq \nu \|Pu\|^2 \; .$$

With this modification, the case $\nu < 0$ may well arise independent of the possible presence of dynamical invariants.

Since $\Delta x(\tau) \in S_P$, we may insert it into the above definition and obtain

$$\hat{\nu}(\tau) = \frac{(\Delta x, A\Delta x)}{\|\Delta x\|^2} \leq \nu \; . \tag{3.2}$$

If we multiply equation (2.3) by Δx from the left, we obtain

$$\begin{aligned} \|\Delta x\|^2 &= \tau(\Delta x, A\Delta x) + (\Delta x, F(x_0)) \\ &= \tau\hat{\nu}(\tau)\|\Delta x\|^2 + (\Delta x, F(x_0)) \\ &\leq \tau\hat{\nu}(\tau)\|\Delta x\|^2 + \|\Delta x\|\|F(x_0)\| \\ &\leq \tau\nu\|\Delta x\|^2 + \|\Delta x\|\|F(x_0)\| \; , \end{aligned}$$

which then leads to the estimates

$$\|\Delta x\| \leq \frac{\|F(x_0)\|}{1 - \hat{\nu}\tau} \leq \frac{\|F(x_0)\|}{1 - \nu\tau} \; . \tag{3.3}$$

Moreover, since

$$\Delta x(\tau) = F(x_0) + \mathcal{O}(\tau) \; ,$$

we also have

$$\hat{\nu}(0) = \frac{(F(x_0), AF(x_0))}{\|F(x_0)\|^2} \leq \nu \; , \tag{3.4}$$

This quantity can be monitored even before the linear equation (2.3) is actually solved. It plays a key role in the residual reduction process, as shown in the following lemma.

Lemma 3.2 *Let $\hat{\nu}(0) < 0$ as defined in (3.4). Then there exists some $\tau^* > 0$ such that*

$$\|F(x(\tau))\| < \|F(x_0)\| \quad \text{and} \quad \hat{\nu}(\tau) < 0 \quad \text{for all} \quad \tau \in [0, \tau^*[\, .$$

Proof. By differentiating the residual norm with respect to τ we obtain

$$\begin{aligned} \frac{d}{d\tau}\|F(x(\tau))\|^2|_{\tau=0} &= 2(F'(\cdot)^T F(\cdot), \dot{x}(\tau))|_{\tau=0} \\ &= 2(F(x_0), AF(x_0)) = 2\hat{\nu}(0)\|F(x_0)\|^2 < 0 \, . \end{aligned}$$

Since both $F(x(\tau))$ and the norm $\|\cdot\|$ are continuously differentiable, there exists some non–empty interval w.r.t. τ, wherein the residual norm decreases. The proof of the statement for $\hat{\nu}(\tau)$ uses the same kind of argument. ∎

In other words: if at the given starting point x_0 the condition $\hat{\nu}(0) < 0$ is violated, then the pseudo-transient continuation method cannot be expected to reduce the residual norm at all. Recall, however, the discussion at the end of Section 2 which pointed out that residual reduction is not equivalent to canonical norm reduction and therefore to contractivity of the differential equation (1.1). As a consequence, if residual reduction fails to occur, then the fixed point may nevertheless be attractive.

Second order Lipschitz condition. Upon recalling that in affine similar nonlinear problems the domain space and the image space of the mapping transform in the same way, we will characterize the nonlinearity by the following Lipschitz condition

$$\|\big(F'(x) - F'(y)\big)u\| \leq L_2\|x - y\|\|u\| \, . \tag{3.5}$$

Convergence analysis. With these preparations we are now ready to state our main result.

Theorem 3.3 *Notation as in the preceding Lemma 3.1. Assume $A = F'(x_0)$ and denote $L_0 = \|F(x_0)\|$. Let dynamical invariants show up via the properties $F(x) \in S_P$. Assume the one–sided first order Lipschitz condition*

$$(u, Au) \leq \nu\|u\|^2 \quad for \quad u \in S_P, \quad \nu < 0,$$

and the second order Lipschitz condition

$$\|\big(F'(x) - F'(x_0)\big)u\| \leq L_2\|x - x_0\|\|u\| \, .$$

Then the following estimate holds

$$\|F(x(\tau))\| \leq \left(1 + \tfrac{1}{2}\frac{L_0 L_2 \tau^2}{1-\nu\tau}\right) \frac{\|F(x_0)\|}{1-\nu\tau} \; . \tag{3.6}$$

From this, residual monotonicity

$$\|F(x(\tau))\| \leq \|F(x_0)\|$$

is guaranteed for all $\tau \geq 0$ *satisfying the sufficient condition*

$$\nu + (\tfrac{1}{2}L_0 L_2 - \nu^2)\tau \leq 0 \; . \tag{3.7}$$

Moreover, if

$$L_0 L_2 > \nu^2 \; , \tag{3.8}$$

then the theoretically optimal pseudo–timestep is

$$\tau_{opt} = \frac{|\nu|}{L_0 L_2 - \nu^2} \; , \tag{3.9}$$

leading to a residual reduction

$$\|F(x(\tau))\| \leq \left(1 - \tfrac{1}{2}\frac{\nu^2}{L_0 L_2}\right) \|F(x_0)\| < \|F(x_0)\| \; .$$

Proof. We return to the preceding Lemma 3.1. Obviously, the first and the second right hand terms in equation (3.1) are independent. Upon recalling (3.3) for the first term, we immediately recognize that, in order to be able to prove residual reduction, we necessarily need the condition $\nu < 0$, which means $\nu = -|\nu|$ throughout the proof. For the second term we may estimate, again recalling (3.5) and (3.3),

$$\begin{aligned}
&\int_{\sigma=0}^{\tau} \| \left(F'(x(\sigma)) - F'(x_0)\right) (I - \sigma A)^{-2} F(x_0) \| d\sigma \\
&\qquad \leq L_2 \int_{\sigma=0}^{\tau} \|x(\sigma) - x_0\| \|(I - \sigma A)^{-2} F(x_0)\| d\sigma \\
&\qquad \leq L_2 \int_{\sigma=0}^{\tau} \frac{\sigma \|F(x_0)\|^2}{(1-\sigma\nu)^3} d\sigma \\
&\qquad = \tfrac{1}{2} L_2 \|F(x_0)\|^2 \tau^2 (1-\nu\tau)^{-2} \; .
\end{aligned}$$

Combination of the two estimates then directly confirms (3.6), which we here write as

$$\|F(x(\tau))\| \leq \alpha(\tau)\|F(x_0)\| \ ,$$

in terms of

$$\alpha(\tau) = \left(1 - \nu\tau + \tfrac{1}{2}L_0L_2\tau^2\right)/(1-\nu\tau)^2 \ .$$

Upon requiring $\alpha(\tau) \leq 1$, we obtain the equivalent sufficient condition (3.7). Finally, in order to find the optimal residual reduction, a short calculation shows that

$$\dot{\alpha}(\tau) = \left(\nu + \frac{L_0L_2\tau}{1-\nu\tau}\right)/(1-\nu\tau)^2 \ .$$

An interior minimum can arise only for $\dot{\alpha}(\tau) = 0$, which is equivalent to (3.9) under the condition (3.8). Insertion of τ_{opt} into the expression for $\alpha(\tau)$ then completes the proof. ■

From the above condition (3.7) we may conclude: if

$$\nu + \tfrac{1}{2}\sqrt{2L_0L_2} \leq 0 \ ;$$

then τ is *unbounded* for local continuation. Apart from a different prefactor, this turns out to be the residual oriented analog of the error oriented condition given in [2]. If

$$\nu + \tfrac{1}{2}\sqrt{2L_0L_2} > 0 \ , \tag{3.10}$$

then the pseudo–timestep is bounded according to

$$\tau \leq \frac{|\nu|}{\frac{1}{2}L_0L_2 - |\nu|^2} \ .$$

Note that condition (3.8) is less restrictive than (3.10) so that either unbounded or bounded optimal timesteps may occur.

Adaptive pseudo–timestep strategy. On the basis of the above convergence analysis we now want to derive an adaptive strategy for the selection of the pseudo–timestep. To do so, we apply the fundamental scheme due to [1] which has been successful so many times. In our context this means to replace the unknown theoretical Lipschitz constants L_2 and ν by carefully selected *computational estimates* $[L_2]$ and $[\nu]$ and to insert them into the formula for the theoretical optimal pseudo–timestep

$$\tau_{opt} = \frac{|\nu|}{L_0L_2 - \nu^2} \ .$$

This expression can be rewritten in implicit form as

$$\tau_{opt} = \frac{|\nu|(1 - \nu\tau_{opt})}{L_0 L_2} \ . \tag{3.11}$$

Remark 3.4 In passing we note that from this representation we may roughly obtain

$$\tau_{opt} \sim \frac{1}{L_0} = \frac{1}{\|F(x_0)\|} \ .$$

This relation gives some justification for a quite popular *heuristic strategy*: new timesteps are proposed on the basis of successful old ones via

$$\tau_{new} = \frac{\|F(x_{old})\|}{\|F(x_{new})\|} \ \tau_{old} \ . \tag{3.12}$$

For reference see, e.g., the paper by Kelley and Keyes [6], where also a whole class of further heuristics is quoted.

In order to exploit the structure of (3.11) in detail, we recall the relation (3.3) and replace τ_{opt} by the upper bound

$$\bar{\tau}_{opt} = \frac{|\nu|}{L_2 \|\Delta x(\tau)\|} \geq \tau_{opt} \ .$$

So we are left with the task of identifying cheap computational estimates $[\nu] \leq \nu < 0, [L_2] \leq L_2$. Once this is achieved, we can compute the corresponding pseudo–timestep

$$[\tau_{opt}] = \frac{|[\nu]|}{[L_2]\|\Delta x(\tau)\|} \geq \bar{\tau}_{opt} \geq \tau_{opt} \ . \tag{3.13}$$

As for the estimation of ν, we may exploit (3.2) in a double way. Whenever

$$\|\Delta x(\tau)\| \geq \|F(x_0)\| \ ,$$

then we know that $\nu \geq [\nu] \geq 0$ is guaranteed: in this situation we should either terminate the iteration or continue without requiring residual reduction (a rather popular heuristics). Otherwise we may recognize that

$$[\nu]\tau = \hat{\nu}(\tau)\tau = \tau \frac{(\Delta x, A\Delta x)}{\|\Delta x\|^2} = \frac{(\Delta x, \Delta x - F(x_0))}{\|\Delta x\|^2} \leq \nu\tau \tag{3.14}$$

gives us a quite cheap estimation formula for ν. As for the estimation of L_2, we may rearrange terms in the proof of Theorem 3.3 to obtain

$$\begin{aligned}\|F(x(\tau)) - \Delta x(\tau)\| &\leq \int_{\sigma=0}^{\tau} \| \left(F'(x(\sigma)) - F'(x_0)\right) (I - \sigma A)^{-2} F(x_0)\| d\sigma \\ &\leq L_2 \int_{\sigma=0}^{\tau} \|x(\sigma) - x_0\| \|(I - \sigma A)^{-1} \Delta x(\sigma)\| d\sigma \ .\end{aligned}$$

If we approximate the integral by the trapezoidal rule, we arrive at

$$\begin{aligned}\|F(x(\tau)) - \Delta x(\tau)\| &\leq \tfrac{1}{2} L_2 \|\Delta x(\tau)\|^2 \tau^2 / (1 - \nu\tau) + \mathcal{O}(\tau^4) \\ &\leq \tfrac{1}{2} L_2 \|\Delta x(\tau)\|^2 \tau^2 + \mathcal{O}(\tau^4) \ .\end{aligned}$$

Note that already the approximation term, ignoring the $\mathcal{O}(\tau^4)$ term, gives rise to the upper bound

$$\|F(x(\tau)) - \Delta x(\tau)\| \leq \tfrac{1}{2} L_2 \|F(x_0)\|^2 \tau^2 / (1 - \nu\tau)^2 \ ,$$

which is the basis of the derivation of Theorem 3.3. Hence, we may well regard

$$[L_2] = \frac{2\|F(x(\tau)) - \Delta x\|}{\tau^2 \|\Delta x\|^2} \leq L_2 + \mathcal{O}(\tau^2)$$

as a suitable computational estimate for L_2. Upon collecting all above estimates and inserting them into (3.13), we arrive at the following pseudo–timestep suggestion

$$[\tau_{opt}] = \frac{|(\Delta x(\tau), F(x_0) - \Delta x(\tau))|}{2\|\Delta x(\tau)\| \|F(x(\tau)) - \Delta x(\tau)\|} \ \tau \ .$$

On this basis, an *adaptive τ–strategy* can be realized in the two usual modes:

- *correction* strategy: if the residual norm does *not* decrease from x_0 to $x(\tau)$, then the actual stepsize τ should be replaced by $[\tau_{opt}] < \tau$;
- *prediction* strategy: if the residual norm decreases, then the next continuatin step is started with the trial value $[\tau_{opt}]$.

Finally, note that the above strategy will terminate, if the steady state to be computed is not attractive in the residual sense, which means that $[\nu] \geq 0$. For $[\nu] \to 0^-$, the suggested stepsize behaves like $[\tau_{opt}] \to 0^+$ – as to be reasonably expected.

4 Pseudo–transient Continuation with Jacobian Approximations

Throughout this section, we consider the case that we only have a Jacobian approximation $A \approx F'(x_0)$ at hand. We still assume that we solve the linear system (2.3) by a direct (possibly sparse) elimination method – for the iterative variant see the subsequent Section 5. In order to preserve the *dynamical invariants*, we restrict the possible approximations such that $P^\perp F(x) = 0$ implies $P^\perp A = 0$. This property will hold, for example, if A is a Jacobian matrix that has been kept over several steps. In order to measure the Jacobian approximation error

$$\delta A = A - F'(x_0)$$

we define an upper bound

$$\|\delta A\, u\| \le \delta|\nu|\|u\|, \quad for \quad u \in S_P\ .$$

Convergence analysis. With this additional notation we are ready to modify the convergence theory of the preceding section.

Theorem 4.1 *Notation as in Lemma 3.1 and Theorem 3.3, but here with $A \approx F'(x_0)$. Let dynamical invariants show up via $F(x) \in S_P$. Assume the approximate one–sided first order Lipschitz condition*

$$(u, Au) \le \nu\|u\|^2, \quad \nu < 0, \quad for \quad u \in S_P, \tag{4.1}$$

and the second order Lipschitz condition (3.5). Let the Jacobian approximation satisfy

$$\|\,(A - F'(x_0))\,u\| \le \delta|\nu|\|u\|, \quad \delta < 1 \quad for \quad u \in S_P\ .$$

Then the following estimate holds

$$\|F(x(\tau))\| \le \left(1 + \delta|\nu|\tau + \tfrac{1}{2}\frac{L_0 L_2 \tau^2}{1-\nu\tau}\right)\frac{\|F(x_0)\|}{1-\nu\tau}\ . \tag{4.2}$$

From this, residual monotonicity

$$\|F(x(\tau))\| \le \|F(x_0)\|$$

holds for all $\tau \geq 0$ *satisfying the sufficient condition*

$$(1-\delta)\nu + \tfrac{1}{2}\frac{L_0 L_2 \tau}{1-\nu\tau} \leq 0\ . \tag{4.3}$$

Moreover, if

$$L_0 L_2 > (1-\delta)\nu^2\ , \tag{4.4}$$

then the theoretically optimal pseudo–timestep is

$$\tau_{opt} = \frac{(1-\delta)|\nu|}{L_0 L_2 - (1-\delta)\nu^2}\ , \tag{4.5}$$

leading to a residual reduction

$$\|F(x(\tau))\| \leq \left(1 - \tfrac{1}{2}(1-\delta)^2 \frac{\nu^2}{L_0 L_2}\right) \|F(x_0)\| < \|F(x_0)\|\ . \tag{4.6}$$

Proof. Lemma 3.1 also holds for $A \neq F'(x_0)$. Hence, the proof of this theorem is just a modification of the proof of the preceding Theorem 3.3. The first right hand term in equation (3.1) can be bounded as before. For the second term we obtain here

$$\begin{aligned} \| \int_{\sigma=0}^{\tau} \left(F'(x(\sigma)) - A\right) (I - \sigma A)^{-2} F(x_0) d\sigma \| \\ \leq \quad \tfrac{1}{2} L_2 \tau^2 |\nu| \|F(x_0)\|^2 (1-\tau\nu)^{-2} \\ + \| \int_{\sigma=0}^{\tau} \left(F'(x_0) - A\right) (I - \sigma A)^{-2} F(x_0) d\sigma \|\ . \end{aligned}$$

The above second term can be treated further observing that

$$\begin{aligned} &\| \int_{\sigma=0}^{\tau} \left(F'(x_0) - A\right) (I - \sigma A)^{-2} F(x_0) d\sigma \| = \\ &= \| \left(F'(x_0) - A\right) \int_{\sigma=0}^{\tau} \dot{x}(\sigma) d\sigma \| = \| \left(F'(x_0) - A\right) (x(\tau) - x_0) \| \\ &\leq \delta |\nu| \tau \|\Delta x(\tau)\| \leq \delta |\nu| \tau \frac{\|F(x_0)\|}{1-\nu\tau}\ , \end{aligned}$$

which confirms (4.2).

In order to analyze residual monotonicity, we rewrite (4.2) in the form

$$F(x(\tau))\| \leq \alpha(\tau)\|F(x_0)\|$$

with

$$\alpha(\tau) = \left((1-\delta\nu\tau)(1-\nu\tau) + \tfrac{1}{2}L_0L_2\tau^2\right)/(1-\nu\tau)^2 \ .$$

The requirement $\alpha \leq 1$ then leads to the sufficient condition (4.3). The optimal pseudo–timestep follows from

$$\dot{\alpha}(\tau) = \left((1-\delta)\nu + \frac{L_0L_2\tau}{1-\nu\tau}\right)/(1-\nu\tau)^2 = 0 \ ,$$

which is equivalent to (4.5) under the condition (4.4). Finally, insertion of τ_{opt} into $\alpha(\tau)$ yields (4.6), which completes the proof. ∎

Note that whenever

$$\nu + \sqrt{\frac{L_0L_2}{2(1-\delta)}} \leq 0 \ ,$$

then τ is *unbounded*. Otherwise the pseudo–timestep is bounded according to

$$\tau \leq \frac{(1-\delta)|\nu|}{\frac{1}{2}L_0L_2 - (1-\delta)\nu^2} \ .$$

Again the stepsize vanishes whenever $\nu \to 0^-$.

Adaptive pseudo–timestep strategy. We proceed along the same lines as in the exact pseudo–transient continuation in Section 3. For convenience, we repeat the theoretical optimal pseudo–timestep

$$\tau_{opt} = \frac{(1-\delta)|\nu|}{L_0L_2 - (1-\delta)\nu^2} = \frac{|\nu|}{L_0L_2/(1-\delta) - \nu^2} \ .$$

The above expression can be rewritten in implicit form as

$$\tau_{opt} = \frac{(1-\delta)|\nu|(1-\nu\tau_{opt})}{L_0L_2} \ .$$

Note that also this representation gives some rough justification for the heuristic strategy (3.12). Next we introduce the theoretical upper bound

$$\bar{\tau}_{opt} = \frac{(1-\delta)|\nu|}{L_2\|\Delta x(\tau)\|} \geq \tau_{opt} \tag{4.7}$$

and its corresponding formal estimate

$$[\tau_{opt}] = \frac{(1-[\delta])|[\nu]|}{[L_2]\|\Delta x(\tau)\|} \geq \bar{\tau}_{opt} \geq \tau_{opt} \ .$$

In contrast to the exact Jacobian case, two conditions are required here:

$$\nu = -|\nu| < 0 \quad \textit{and} \quad \delta < 1 \ ,$$

or, equivalently,

$$(1-\delta)\nu < 0 \ .$$

In the fortunate case that $\|\delta A\| \ll |\nu|$, which is equivalent to $\delta \ll 1$, we may just use the adaptive stepsize control as described in the preceding Section 3. If, however, the Jacobian approximation errors cannot be neglected, we have to take this error into account when estimating the theoretical quantities needed in (4.7).

To begin with, we will need to replace the quantity $(1-\delta)\nu$ by some lower bound

$$[\nu] + [\|\delta A\|] \leq \nu + \delta|\nu| = (1-\delta)\nu \ .$$

Consequently, the condition

$$[\nu] + [\|\delta A\|] < 0 \tag{4.8}$$

will be necessary for the continuation process to converge to the steady state solution. For the actual realization, two typical cases need to be discussed.

In the *first* case, the Jacobian error $\|\delta A\|$ can only be estimated via theoretical upper bounds. For example, when the Jacobian has been "kept" from a previous continuation step at $x(\hat{\tau})$, the error can be estimated via

$$\|\delta A\| = \|F'(x_0) - F'(x(\hat{\tau}))\| \leq L_2\|x_0 - x(\hat{\tau})\| \ .$$

Since the term $\|x_0 - x(\hat{\tau})\|$ is computationally available, a value for $[\|\delta A\|]$ can be assigned, once an estimate $[L_2]$ is given — which can be also kept from the continuation history. If, in addition, we compute an updated estimate $[\nu]$ from (3.14), Jacobian regeneration can be controlled such that (4.8) holds throughout the continuation process. On this basis, we may realize an adaptive pseudo–transient continuation method.

In the *second* case, the Jacobian matrix $F'(x_0)$ is actually known, but not fully used to save linear algebra work in the numerical solution of the pseudo–transient continuation equation — a strategy known as *sparsing*. In this

case, we may nevertheless use the full Jacobian information for the purpose of estimating the theoretical quantities needed in (4.7) in a clean way. By comparison with (4.1) we observe that

$$(u, F'(x_0)u) = (u, Au) - (u, \delta Au) \leq (1-\delta)\nu\|u\|^2 \ .$$

From this we are immediately led to the computational lower bound

$$[(1-\delta)\nu] := \frac{(\Delta x, F'(x_0)\Delta x)}{\|\Delta x\|^2} \leq (1-\delta)\nu \ , \tag{4.9}$$

which is comparable to the estimate (3.14). Note that here the condition $\|\Delta x(\tau)\| < \|F(x_0)\|$ is still necessary to guarantee at least $[\nu] < 0$. Moreover, in the light of (4.9), the condition (4.8) is understood to be necessary for the steady state to be attractive (in the residual sense, of course).

As for the estimation of L_2, we rearrange terms in the proof of Theorem 4.1 and apply again the trapezoidal rule so that

$$\begin{aligned} &\|F(x(\tau)) - (F(x_0) + \tau F'(x_0)\Delta x(\tau))\,\| \\ &\leq \int_{\sigma=0}^{\tau} \|\,(F'(x(\sigma)) - F'(x_0))\,\dot{x}(\sigma)\|d\sigma \\ &\leq \tfrac{1}{2}L_2\|\Delta x(\tau)\|^2\tau^2 + \mathcal{O}(\tau^4) \ . \end{aligned}$$

If we, as in the exact Jacobian case, once more ignore the $\mathcal{O}(\tau^4)$ term, we arrive at the modified estimate

$$[L_2] = \frac{2\|F(x(\tau)) - (F(x_0) + \tau F'(x_0)\Delta x(\tau))\,\|}{\tau^2\|\Delta x\|^2} \leq L_2 + \mathcal{O}(\tau^2)$$

and finally at the suggested pseudo–timestep

$$[\tau_{opt}] = \frac{|(\Delta x(\tau), F'(x_0)\Delta x(\tau))|}{2\|\Delta x(\tau)\|\|F(x(\tau)) - (F(x_0) + \tau F'(x_0)\Delta x(\tau))\,\|}\ \tau^2 \ .$$

Again this gives rise to a *correction* and a *prediction* strategy for an adaptive continuation process.

5 Inexact pseudo–transient continuation

Suppose the linear system (2.3) is so large that we cannot but solve it *iteratively* ($i = 0, 1, \ldots$):

$$(I - \tau A)\delta x_i = F(x_0) - r_i, \quad x_i(\tau) = x_0 + \tau \delta x_i \ . \tag{5.1}$$

Herein r_i represents the iterative linear residual, δx_i the corresponding inexact correction, and $x_i(\tau)$ the approximate homotopy path instead of the exact $x(\tau)$. To start the iteration, let $x_0(\tau) = x_0$ so that $\delta x_0 = 0$ and $r_0 = F(x_0)$. In the context of discretized PDEs the matrix A is typically the exact sparse Jacobian.

If we want to *minimize the residuals* within each iterative step, we are directly led to GMRES [7]. In terms of the Euclidean norm $\|\cdot\|$ we define the approximation quantities

$$\eta_i := \frac{\|r_i\|}{\|F(x_0)\|} < 1 \quad for \quad i = 1, 2, \ldots \ .$$

Recall that GMRES assures $\eta_{i+1} \leq \eta_i$, in the generic case even $\eta_{i+1} < \eta_i$. Moreover, due to the residual minimization property and $r_0 = F(x_0)$, we have

$$\|F(x_0) - r_i\|^2 = (1 - \eta_i^2)\|F(x_0)\|^2 \ .$$

In the present context of pseudo–transient continuation, we may additionally observe that GMRES realizes the special structure

$$\delta x_i(\tau) = V_i z_i(\tau) \quad and \quad H_i(\tau) = (I_i, 0)^T + \tau \hat{H}_i \ .$$

Herein V_i is just the orthonormal basis of the Krylov space $\mathcal{K}_i(r_0, A)$ and $\hat{H}_i$ is a Hessenberg matrix like $H_i(\tau)$, but also independent of τ. On this basis, dynamical invariants can be shown to be treated correctly throughout the iteration (proof omitted here). The special structure also permits computational savings when the same system is solved for different pseudo–timesteps.

Convergence analysis. As in Sections 3 and 4 we first analyze the convergence behavior theoretically as a basis for the subsequent derivation of an adaptive algorithm which here will include the matching of inner and outer iteration. Before we can do so, we need to modify the auxiliary Lemma 3.1.

Lemma 5.1 *Notation as in Lemma 3.1 with $A \approx F'(x_0)$. Then the residual along the approximate homotopy path $x_i(\tau)$ starting at x_0 satisfies*

$$F(x(\tau)) - r_i = (I - \tau A)^{-1} (F(x_0) - r_i)$$

$$+ \int_{\sigma=0}^{\tau} (F'(x_i(\sigma)) - A) (I - \sigma A)^{-2} (F(x_0) - r_i) \, d\sigma \ .$$

Proof. The proof is an elementary modification of the proof of Lemma 3.1. If we differentiate the homotopy (5.1) with respect to τ, we now obtain

$$\dot{x}_i(\tau) = (I - \tau A)^{-2} (F(x_0) - r_i) \ .$$

Further details can be omitted. ∎

Theorem 5.2 *Notation as in the preceding Lemma 5.1. Let $A = F'(x_0)$ and $\tilde{L}_0 = \sqrt{1 - \eta_i^2} \|F(x_0)\|$. Assume that dynamical invariants show up via the properties $F(x) \in S_P$. Then, with the Lipschitz conditions*

$$(u, Au) \leq \nu \|u\|^2, \quad \nu < 0, \quad for \quad u \in S_P$$

and

$$\| \big(F'(x) - F'(x_0)\big) u \| \leq L_2 \|x - x_0\| \|u\| \ ,$$

the estimates

$$\|F(x(\tau)) - r_i\| \leq \left(1 + \tfrac{1}{2} \frac{\tilde{L}_0 L_2 \tau^2}{1 - \nu\tau}\right) \frac{\|F(x_0) - r_i\|}{1 - \nu\tau}$$

and

$$\|F(x(\tau))\| \leq \left(\eta_i + \frac{\sqrt{1 - \eta_i^2}}{1 - \nu\tau} \left(1 + \tfrac{1}{2} \frac{\tilde{L}_0 L_2 \tau^2}{1 - \nu\tau}\right)\right) \|F(x_0)\| \ .$$

hold. Let

$$s(\eta_i) := \sqrt{\frac{1 - \eta_i}{1 + \eta_i}} > \tfrac{1}{2} \quad or, equivalently \quad \eta_i < \tfrac{3}{5} \ . \tag{5.2}$$

Then residual monotonicity

$$\|F(x(\tau))\| \leq \|F(x_0)\|$$

is guaranteed for all $\tau \geq 0$ *satisfying the sufficient condition*

$$1 - s(\eta_i) + (2s(\eta_i) - 1)\nu\tau + \left(\tfrac{1}{2}\tilde{L}_0 L_2 - s(\eta_i)\nu^2\right)\tau^2 \leq 0 \; . \tag{5.3}$$

Assume further that

$$\tfrac{4}{5} L_0 L_2 > \nu^2 \tag{5.4}$$

and that GMRES *has been continued until*

$$\eta_i + \sqrt{1 - \eta_i^2} < 1 + \tfrac{1}{2}\frac{\nu^2}{L_0 L_2} \; . \tag{5.5}$$

Then the theoretically optimal pseudo–timestep is

$$\tau_{opt} = \frac{|\nu|}{\tilde{L}_0 L_2 - \nu^2} \; , \tag{5.6}$$

leading to the estimate

$$\|F(x(\tau)) - r_i\| \leq \left(1 - \tfrac{1}{2}\frac{\nu^2}{\tilde{L}_0 L_2}\right)\|F(x_0) - r_i\| < \|F(x_0) - r_i\|$$

and to the residual reduction

$$\|F(x(\tau))\| \leq \left(\eta_i + \sqrt{1 - \eta_i^2} - \tfrac{1}{2}\frac{\nu^2}{L_0 L_2}\right)\|F(x_0)\| < \|F(x_0)\| \; . \tag{5.7}$$

Proof. We return to the preceding Lemma 5.1 and modify the proof of Theorem 3.3 carefully step by step. For example, the second order term may be estimated as

$$\begin{aligned} \int_{\sigma=0}^{\tau} & \| \left(F'(x(\sigma)) - F'(x_0)\right)(I - \sigma A)^{-2}(F(x_0) - r_i)\| d\sigma \\ & \leq \tfrac{1}{2} L_2 \|F(x_0) - r_i\|^2 \tau^2 (1 - \nu\tau)^{-2} \; . \end{aligned}$$

Combination of estimates then directly confirms

$$\|F(x_i(\tau)) - r_i\| \leq \bar{\alpha}_i(\tau)\|F(x_0) - r_i\|$$

in terms of

$$\bar{\alpha}_i(\tau) = \left(1 - \nu\tau + \tfrac{1}{2}\tilde{L}_0 L_2 \tau^2\right)/(1 - \nu\tau)^2 \; ,$$

from which we obtain

$$\|F(x_i(\tau))\| \leq \alpha_i(\tau)\|F(x_0)\|$$

with

$$\alpha_i(\tau) = \eta_i + \sqrt{1-\eta_i^2}\bar{\alpha}_i(\tau) \ .$$

Upon requiring $\alpha(\tau) \leq 1$, we have

$$\eta_i + \sqrt{1-\eta_i^2}\bar{\alpha}_i(\tau) \leq 1 \ ,$$

which is equivalent to

$$\bar{\alpha}_i(\tau) \leq s(\eta_i) \leq 1 \ . \tag{5.8}$$

From this, we immediately verify the sufficient condition (5.3). Note that $2s - 1 > 0$, which is just condition (5.2), is necessary to have at least one negative term in the left hand side of (5.3).

Finally, in order to find the *optimal* residual reduction, a short calculation shows that

$$\dot{\alpha}(\tau) = \sqrt{1-\eta_i^2}\dot{\bar{\alpha}}(\tau) = \frac{\sqrt{1-\eta_i^2}}{(1-\nu\tau)^2}\left(\nu + \frac{\tilde{L}_0 L_2 \tau}{1-\nu\tau}\right) \ .$$

For the interior minimum we require $\dot{\bar{\alpha}}(\tau) = 0$, which is equivalent to (5.6) under the condition (5.4), where

$$\sqrt{1-\eta_i^2} \geq \sqrt{1-(\tfrac{3}{5})^2} = \tfrac{4}{5}$$

has been used. Insertion of τ_{opt} into the expression for $\alpha(\tau)$ then leads to

$$\|F(x_i(\tau)) - r_i\| \leq \left(1 - \tfrac{1}{2}\frac{\nu^2}{\tilde{L}_0 L_2}\right)\|F(x_0) - r_i\|$$

and eventually to (5.7). In order to assure an actual *residual reduction*, condition (5.8) must also hold for τ_{opt}, which confirms the necessary condition (5.5). Note that the scalar function $\eta_i + \sqrt{1-\eta_i^2}$ is monotonically increasing for $\eta_i < \frac{1}{2}\sqrt{2} \approx 0.7$, hence also for $\eta_i < \frac{3}{5} = 0.6$. Therefore GMRES may be just continued until the relation (5.5) is satisfied. This completes the proof. ∎

Adaptive pseudo–timestep strategy. We follow the lines of the derivation for the exact pseudo–transient continuation in Section 3. To start with, we repeat the expression

$$\tau_{opt} = \frac{|\nu|}{\tilde{L}_0 L_2 - \nu^2} \ ,$$

which can be rewritten in implicit form as

$$\tau_{opt} = \frac{|\nu|(1 - \nu\tau_{opt})}{\tilde{L}_0 L_2} \ .$$

Recall now that

$$\|\delta x_i(\tau)\| = \|(I - \tau A)^{-1}(F(x_0) - r_i)\| \leq \frac{\|F(x_0) - r_i\|}{1 - \nu\tau} = \frac{\tilde{L}_0}{1 - \nu\tau} \ , \qquad (5.9)$$

which directly implies

$$\bar{\tau}_{opt} = \frac{|\nu|}{L_2 \|\delta x_i(\tau)\|} \geq \tau_{opt} \ .$$

So we need to compute the pseudo–timestep

$$[\tau_{opt}] = \frac{|[\nu]|}{[L_2]\|\delta x_i(\tau)\|} \geq \bar{\tau}_{opt} \geq \tau_{opt} \ . \qquad (5.10)$$

in terms of the appropriate estimates of the unknown theoretical quantities ν, L_2.
As for the estimation of ν, we exploit (5.9). Whenever

$$\|\delta x_i(\tau)\| \geq \|F(x_0) - r_i\| \ ,$$

then we know that $\nu \geq 0$ is guaranteed and the iteration must be terminated. Moreover, the relation

$$[\nu]\tau = \tau \frac{(\delta x_i, A\delta x_i)}{\|\delta x_i\|^2} = \frac{(\delta x_i, \delta x_i - F(x_0) + r_i)}{\|\delta x_i\|^2} \leq \nu\tau$$

supplies an estimation formula for ν. As for the estimation of L_2, we revisit Lemma 5.1 to obtain

$$\begin{aligned}
&\|F(x_i(\tau)) - r_i - \delta x_i(\tau)\| \\
&\quad \leq \int_{\sigma=0}^{\tau} \| \left(F'(x_i(\sigma)) - F'(x_0)\right) (I - \sigma A)^{-2} \left(F(x_0) - r_i\right) \| d\sigma \\
&\quad \leq L_2 \int_{\sigma=0}^{\tau} \|x_i(\sigma) - x_0\| \|(I - \sigma A)^{-1} \delta x_i(\sigma)\| d\sigma \\
&\quad \leq \tfrac{1}{2} L_2 \tau^2 \frac{\tilde{L}_0^2}{(1 - \nu\tau)^2} \; .
\end{aligned}$$

If we approximate the above integral by the trapezoidal rule (*before* using the final estimate), we arrive at

$$\begin{aligned}
\|F(x_i(\tau)) - r_i - \delta x_i(\tau)\| &\leq \tfrac{1}{2} L_2 \|\delta x_i(\tau)\|^2 \tau^2 / (1 - \nu\tau) + \mathcal{O}(\tau^4) \\
&\leq \tfrac{1}{2} L_2 \tau^2 \|\delta x_i(\tau)\|^2 + \mathcal{O}(\tau^4) \; .
\end{aligned}$$

Already the first right hand term gives rise to the above upper bound — compare (5.9). Hence, as in Section 3, we will pick

$$[L_2] = \frac{2\|F(x_i(\tau)) - r_i - \delta x_i(\tau)\|}{\tau^2 \|\delta x_i(\tau)\|^2} \leq L_2 + \mathcal{O}(\tau^2)$$

as computational estimate for L_2. Upon inserting the two derived estimates into (5.10), we arrive at the pseudo–timestep estimate

$$[\tau_{opt}] = \frac{|(\delta x_i(\tau), F(x_0) - r_i - \delta x_i(\tau))|}{2\|\delta x_i(\tau)\| \|F(x_i(\tau)) - r_i - \delta x_i(\tau)\|} \; \tau \; .$$

On this basis, an *adaptive τ–strategy* can again be realized as in the case of the *exact* pseudo–transient continuation method.

Finally, we want to mention that the iterative version of the pseudo–transient continuation method still works in the case of *unbounded* timestep. To see this, just rewrite (5.1) in the form

$$\left(\frac{1}{\tau} I - A\right) (x_i(\tau) - x_0) = F(x_0) - r_i \; .$$

Herein $\tau \to \infty$ is possible leaving $x_i(\tau) - x_0$ well–defined even in the presence of singular Jacobian A caused by *dynamical invariants*: this is due to the fact that GMRES (like any Krylov solver) keeps the nullspace components of the solution unchanged, so that $P^{\perp}(x_i(\infty) - x_0) = 0$ is guaranteed throughout the iteration.

Preconditioning. I we modify the nonlinear system by means of some nonsingujlar matrix M from the left as

$$MF(x) = 0 \ ,$$

then GMRES will have to work on the preconditioned residuals Mr_i and adaptivity must be based on norms $\|M \cdot \|$. Preconditioning from the right will just influence the convergence speed of GMRES without changing the above derived adaptivity devices.

Derivative–free realization. Sometimes the inexact pseudo–continuation method is realized in a derivative–free variant using the first order approximation

$$A\delta x \approx F(x + \delta x) - F(x) \ .$$

To study this variant, additional errors δA for the second order discretization error need to be taken into account requiring a proper combination of the results of Section 4 and of this section.

Conclusion

The paper presents a convergence analysis as a framework for adaptive pseudo–transient continuation algorithms. Special focus is on large scale systems such as discretized PDEs. The suggested algorithms include the cases when only an approximate (instead of an exact) Jacobian matrix is available or when the arising linear systems are solved by an iterative method (like GMRES). On this basis, numerical experiments with realistic nonlinear systems can be carried out. Preconditioning will turn out to be indispensable.

References

[1] P. Deuflhard. A stepsize control for continuation methods and its special application to multiple shooting techniques. *Numer. Math.*, 33:115–146, 1979.

[2] P. Deuflhard. Uniqueness theorems for stiff ode initial value problems. In D.F. Griffiths and G.A. Watson, editors, *Proceedings 13th Biennial Conference on Numerical Analysis 1989*, pages 74–205, Harlow, Essex, UK, 1990. Longman Scientific & Technical.

[3] P. Deuflhard and F. Bornemann. *Scientific Computing with Ordinary Differential Equations*. Springer, New York, 2002.

[4] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I, Nonstiff Problems*. Springer Verlag, Berlin, Heidelberg, New York, Tokyo, 1987.

[5] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II. Stiff and Differential–Algebraic Problems*. Springer–Verlag, Berlin, Heidelberg, New York, 2nd edition, 1996.

[6] C. T. Kelley and D. E. Keyes. Convergence analysis of pseudo–transient continuation. *SIAM J. Numer. Anal.*, 35:508–523, 1998.

[7] Y. Saad. *Iterative methods for sparse linear systems*. PWS publishing, New York, 1996.